NATIONAL BUREAU OF STANDARDS REPORT

1129

ON CERTAIN ESTIMATORS

BASED ON LARGE SAMPLES OF EXTREMES

By

Julius Lieblein

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

U. S. DEPARTMENT OF COMMERCE
Charles Sawyer, Secretary

NATIONAL BUREAU OF STANDARDS
E. U. Condon, Director

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

1. ELECTRICITY. Resistance Measurements. Inductance and Capacitance. Electrical Instruments. Magnetic Measurements. Electrochemistry.
2. OPTICS AND METROLOGY. Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Gage.
3. HEAT AND POWER. Temperature Measurements. Thermodynamics. Cryogenics. Engines and Lubrication. Engine Fuels. Combustion.
4. ATOMIC AND RADIATION PHYSICS. Spectroscopy. Radiometry. Mass Spectrometry. Physical Electronics. Electron Physics. Atomic Physics. Neutron Measurements. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Instruments Branch.
5. CHEMISTRY. Organic Coatings. Surrace Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.
6. MECHANICS. Sound. Mechanical Instruments. Aerodynamics. Engineering Mechanics. Hydraulics. Mass. Capacity, Density, and Fluid Meters.
7. ORGANIC AND FIBROUS MATERIALS. Rubber. Textiles. Paper. Leather. Testing and Specifications. Organic Plastics. Dental Research.
8. METALLURGY. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.
9. MINERAL PRODUCTS. Porcelain and Pottery. Glass. Refractories. Enameled Metals. Building Stone. Concreting Materials. Constitution and Microstructure. Chemistry of Mineral Products.
10. BUILDING TECHNOLOGY. Structural Engineering. Fire Protection. Heating and Air Conditioning. Exterior and Interior Coverings. Codes and Specifications.
11. APPLIED MATHEMATICS. Numerical Analysis. Computation. Statistical Engineering. Machine Development.
12. ELECTRONICS. Engineering Electronics. Electron Tubes. Electronic Computers.
13. ORDNANCE DEVELOPMENT. Mechanical Research and Development. Electromechanical Fuzes. Technical Services. Missile Fuzing Research. Missile Fuzing Development. Projectile Fuzes. Ordnance Components. Ordnance Tests. Ordnance Research.
14. RADIO PROPAGATION. Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Frequency Utilization Research. Tropospheric Propagation Research. High Frequency Standards. Microwave Standards.
15. MISSILE DEVELOPMENT. Missile Engineering. Missile Dynamics. Missile Intelligence. Missile Instrumentation. Technical Services.

NBS PROJECT

1103-21-5106

August 15, 1951

NBS REPORT

1129

ON CERTAIN ESTIMATORS

BASED ON LARGE SAMPLES OF EXTREMES

By

Julius Lieblein

Research supported by the
National Advisory Committee
for Aeronautics

FOREWORD

The methods of analysis of extreme-value data developed in the present report will be useful in the economical handling of large amounts of extreme gust-load data for airplanes in flight. These methods will in many cases give better results than procedures used previously which required a much larger amount of calculation.

The development of these newer methods is one phase of a project aimed at the improved application of the theory of extreme values to the analysis of gust loads of airplanes. This research, carried out by Mr. Julius Lieblein under the general supervision of Dr. Churchill Eisenhart, Chief of the Statistical Engineering Laboratory, was supported by the National Advisory Committee for Aeronautics. The Statistical Engineering Laboratory is Section 11.3 of the National Applied Mathematics Laboratories (Division 11, National Bureau of Standards), and is concerned with the development and application of modern statistical methods in the physical sciences and engineering.

J. H. Curtiss
Chief, National Applied
Mathematics Laboratories

E. U. Condon
Director
National Bureau of Standards

ON CERTAIN ESTIMATORS BASED ON LARGE SAMPLES OF EXTREMES

Contents

ON CERTAIN ESTIMATORS BASED ON LARGE SAMPLES OF EXTREMES

Julius Lieblein*

I. SUMMARY

Statistical techniques are developed and applied for the economical handling of large masses of extreme gust-load data. These methods are especially well adapted for punch-card equipment, requiring essentially just a counting sorter. The efficiencies of the proposed methods when used with large samples are in some cases superior to the efficiencies of certain methods proposed by E. J. Gumbel which have been used in an NACA publication (reference 1).

This report is limited to methods applicable to large samples. The principal technique is the use of simple sums and differences of order statistics. If the <u>n</u> values in a sample are arranged in (say) ascending order, they are called <u>order statistics</u>. If the data may be assumed to come from a distribution of extreme values, these order statistics can be used in estimating the parameters of the fitted distribution. This report describes and gives the efficiencies of rapid methods:

(i) for estimating the population mode when the dispersion is known by using 1, 2, 3, 4, or 5 suitably chosen order statistics from the sample of n;

(ii) for estimating the population standard deviation by using 2 or 4 suitable order statistics.

These methods are summarized in Table I; their use is explained in the text of this report, and listed in condensed form under Conclusions.

---

* Statistical Engineering Laboratory, National Bureau of Standards. Preparation of this paper was sponsored by the National Advisory Committee for Aeronautics.

II. INTRODUCTION

The purpose of this report is to contribute some results on the use of order statistics in the analysis of large bodies of extreme gust-load data. Reference 1 indicates that the distribution of extreme values of the form

$$F(x) = \exp[-e^{-\alpha(x-u)}]$$

$$= \text{Probability that an observed value is} \leq x,$$

is applicable to problems of predicting the frequency of encountering very severe gust loads and gust velocities under certain test and operating conditions. Applicability of the extreme-value distribution will therefore be assumed for purposes of this report.

The problem is then to fit an extreme-value distribution to a sample of data in order to provide a basis for analysis and prediction. A method of fitting by estimating the two parameters, u (the mode) and $1/\alpha$ (the scale factor or dispersion), has been proposed by E. J. Gumbel(reference 2, lecture 3, p. 18). As a result of preliminary analysis it appears that this method may have low efficiency relative to the best efficiency possible, efficiency being measured relative to the number of observations necessary to assure a specified degree of precision.

It is therefore desirable to investigate alternative methods of estimation for possible improvement in efficiency.

To avoid ambiguity, it is essential to distinguish clearly between an estimator and an estimate. An estimator of a parameter is a mathematical function of sample values used to estimate the parameter. An estimate is the numerical value of this function obtained by substituting into it the values in an actual sample of observations.

One type of approach has been given by F. Mosteller (reference 3) for large samples from a normal population. If the n sample values are arranged in (say) ascending order:

$$x_1, x_2, \ldots, x_n,$$

they are called order statistics with ranks 1, 2, ...n, respectively. Mosteller has found a method of choosing in advance k of the n ranks, k = 1, 2, ..., 10, such that the correspondingly spaced order statistics serve to build an estimator for the mean of a normal distribution which is up to 95 percent efficient, and an estimator for the standard deviation which is up to 75 percent efficient.

The methods used by Mosteller have been adapted and extended in this report to give results for extreme-value samples analogous to results he obtained for samples from a normal distribution.

The big advantage in the use of order statistics for large samples is that three or fewer of the n order statistics can give as much precision as the use of all n values treated by ordinary methods. For example, for n = 1000, the 200th value in order of size estimates the mode of the

population with greater precision than the average of all 1000 values. For large samples the appropriate sample values can readily be selected from the data by mechanical sorting, provided the data are available in complete detail, before grouping or other processing.

III. SYMBOLS

x	random variable
$F(x)$	cumulative probability function
$f(x)$	density function, derivative of $F(x)$
α	scale parameter of distribution of largest values
β	same as $1/\alpha$
u	mode of distribution of largest values
n, N	number of observations in a sample
$x_1, x_2, \ldots, x_n$	values in a sample of size n arranged in increasing order
k	number of order statistics used in estimating a parameter
$\hat{u}$	any estimator of parameter u; also used for estimator $\bar{x} - C\beta$
$\hat{\beta}$	any estimator of parameter β
$\hat{u}_k$	estimator of parameter u formed from k order statistics
$\hat{\beta}_k$	estimator of parameter β formed from k order statistics
$\hat{u}'$	order statistic estimator of u when β is unknown
$\hat{u}_G$	Gumbel's estimator of u when β is unknown
v_k	unbiased estimator of u formed from k order statistics

$E(T)$ expected value of any estimator T

$\sigma^2(T)$ or $V(T)$ variance of T

$\lambda_i = n_i/n$, ratio of rank of $n_i\underline{^{th}}$ order statistic to sample size n

b_N bias in equidistant method of spacing order statistics in sample of N largest values

B_N bias in Gumbel's plotting method for sample of N largest values

$x_{i,k}$ $i\underline{^{th}}$ order statistic in sample of k values from extreme-value distribution with parameters α, u

$y_{i,k}$ $i\underline{^{th}}$ order statistic in sample of k values from extreme-value distribution with parameters $\alpha = 1$, $u = 0$

C Euler's number, 0.57721566....

c correction factor for unbiased estimate of dispersion parameter

s standard deviation of observed sample values; $s = \sqrt{\Sigma(x_i - \bar{x})^2/n}$

IV. EXTREME-VALUE PARAMETERS

The extreme-value distribution to be fitted to a sample of data has the form

(1) $F(x) \equiv \exp[-e^{-\alpha(x-u)}]$

$= $ Probability that an observed value is $< x$.

This is the cumulative form of the distribution. The frequency, or density, form is given by the derivative

$$f(x) = F'(x) \equiv \alpha \exp[-\alpha(x-u) - e^{-\alpha(x-u)}] .$$

This function has the general shape shown in Figure 1, which is scaled so that $\alpha = 1$ and $u = 0$. The function $F(x)$

represents the area under the curve $f(x)$ to the lef of the ordinate drawn at the point x.

Fitting the extreme-value distribution consist in estimating the parameters α and u from the data in hand, or of estimating one of these parameters when the othe is known from previous data. These parameters have meanings analogous to the standard deviation σ and mean (or mode) μ in the case of the normal distribution

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] .$$

The quantity $1/\alpha$, like σ, is a scale parameter measuring dispersion about the central value; u is a location parameter analogous to μ of the normal distribution, and designates the most probable value or mode. The mean of the extreme-value distribution is

$$(2) \quad E(x) = u + \frac{C}{\alpha} ,$$

where C is Euler's number, 0.57721566... Since it is more natural to deal with $1/\alpha$ rather than α, we put

$$\beta = 1/\alpha$$

and consider β as the parameter to be estimated. β and u will also be referred to as extreme-value, or extremal, parameters.

V. ESTIMATION OF MODE u WHEN β IS KNOWN

Estimation of the mode in the case where the other parameter, β, is known, consists in choosing a specified k of the sample values $x_1, x_2, \ldots, x_n$ and computing their mean:

$$\hat{u}_k = \frac{1}{k} \sum_{i=1}^{k} x_{n_i}$$

where $\hat{u}_k$ is the estimator of u formed with the k order statistics $x_{n_1}, x_{n_2}, \ldots, x_{n_k}$. The problem is how to select the k values $n_1, n_2, \ldots, n_k$.

Three methods of selecting, or spacing, are considered in this report: (i) "optimum" spacing, which gives a more efficient estimator than any other method of spacing; (ii) equidistant spacing, used in E. J. Gumbel's graphical method of analysis (reference 1, Lecture 2, pp. 20-21); and (iii) a method of equating expected values analogous to one advocated by B. F. Kimball (reference 5). These estimators are compared below with each other and with an estimator given by Gumbel which uses the mean of all sample values.

1. <u>"Optimum" spacing.</u>

a. <u>Use of one sample value.</u> From the expression for F(x) in equation (1) we see that the probability that a value is less than the mode, x = u, is

$$F(u) = e^{-1} = 0.36788 \quad .$$

Hence in a sample of, say, 1,000, we would expect the 368th value in order of size to give a good estimate of u. In general, for n large, our first estimator of the mode u is

$$\hat{u}_1 = x_{n_1} = x_{.368n}$$

where

$\hat{u}_1$ denotes an estimator of u using one sample value;

$x_{.368n}$ denotes the ordered sample value whose rank is .368 times sample size n.

The quantity $\hat{u}_1$ is a statistic whose value varies from sample to sample. Its distribution is approximately normal for n large — the larger the n the closer the approximation — with (asymptotically)[1]

(3) $\quad \text{Mean } \hat{u}_1 = E(\hat{u}_1) = u$

(4) $\quad \text{variance of } \hat{u}_1 = \sigma^2(\hat{u}_1) = \dfrac{\lambda(1-\lambda)}{n[f(u)]^2}$

where

$$\lambda = F(u) = \frac{1}{e}$$

$$f(u) = \left.\frac{dF(x)}{dx}\right]_{x=u} = \frac{\alpha}{e} = \frac{1}{\beta e}.$$

---

[1] Unless otherwise indicated, all formulas in this report involving characteristics of estimators such as mean, variance, bias, efficiency, will be understood to hold in an asymptotic sense only, i.e. in the limit as sample size n becomes indefinitely large. But for all finite n, however large, the relationships are to be regarded as merely approximate, the approximation being better (in general) the larger the sample size.

Since the average value of $\hat{u}_1$ in repeated sampling is u, the parameter which is estimated, we say that $\hat{u}_1$ is (asymptotically) an unbiased[2] estimator. The variance $\sigma^2(\hat{u}_1)$ (also denoted by $V(\hat{u}_1)$) is

$$\sigma^2(\hat{u}_1) = \frac{(e-1)\beta^2}{n} .$$

In this relation the variance on the left measures the degree of precision obtainable from a sample of n observations, and the equation shows that precision is inversely proportional to the sample size.

An inequality (due to Cramér and Rao, reference 7) in the theory of statistical estimation shows that under certain general conditions the smallest variance (and hence greatest precision) obtainable with any unbiased estimator of the parameter u, for a sample of size m, is not less than

$$\sigma^2(\hat{u})_{min} = \frac{\beta^2}{m} .$$

From the above it is seen that when $\sigma^2(\hat{u}_1) = \sigma^2(\hat{u})_{min}$,

$$\frac{m}{n} = \frac{1}{e-1} = 0.5820.$$

This means that for a given degree of precision, the theoretically smallest number of observations necessary

[2] The term unbiased in statistics is generally reserved for the case of complete absence of bias in samples of any size, small or large. For simplicity, since the present report deals only with large n, the term is applied also to estimators which are not strictly unbiased for finite n, as long as their bias disappears when n increases without limit.

($m = .5820\ n$) is not much over half the number (n) required by the given estimator $\hat{u}_1$. This condition is true generally — the actual number of observations is always greater than (or in some cases equal to) the theoretically smallest number. The ratio m/n thus has values only from 0 to 1, and may therefore be used as a measure of "efficiency." It is more conveniently calculated from the ratio $\sigma^2(\hat{u})/\sigma^2(\hat{u}_1)$ formed for $m = n$, and we write

$$\mathrm{eff}(\hat{u}_1) = \frac{\sigma^2(\hat{u})_{\min}}{\sigma^2(\hat{u}_1)} = \frac{1}{e-1} = .5820$$

and say that the estimator $\hat{u}_1$ is about 58 percent as efficient as a theoretically most efficient estimator.

Can a more efficient estimator be obtained? If in place of $\hat{u}_1 = x_{n_1}$ above, we use any other ordered sample value

(5) $\hat{u}_1{}' = x_{n_1{}'}$

then Mosteller (reference 3) has shown that

(6) $E(\hat{u}_1{}') = u'$

(7) $$\sigma^2(\hat{u}_1{}') = \frac{\lambda_1(1-\lambda_1)}{n[f(u')]^2}$$

where the rank

$$n_1' = \lambda_1 n \quad , \quad 0 < \lambda_1 < 1 \quad ,$$

and u' is the position of the ordinate which cuts off the area λ_1 (on the left) under the curve $f(x)$.

By varying λ_1 it is found by trial that

$$\lambda_1 = .20$$

gives the smallest variance and therefore the greatest efficiency, 64.8 percent. Hence, theoretically,

$$\hat{u}_1' = x_{.20n}$$

gives a "best" or "optimum" estimator when a single sample value is used.

However, if $\hat{u}_1'$ is used to estimate u, a correction must be added, since by equation (6), $\hat{u}_1'$ estimates not u but u' and is therefore <u>biased</u>. The value of the bias is defined as the amount by which an estimator of a parameter over or underestimates it (<u>on the average</u>), the bias being positive or negative accordingly. Using the fact that the area under f(x) up to the ordinate at x is F(x), we have from the definition of u', equation (1), and the definition of λ_1,

$$(8) \quad F(u') \equiv \exp[-e^{-\alpha(u'-u)}] = \lambda_1 \quad ,$$

whence

$$(9) \quad u' = u + \frac{1}{\alpha}[-\ln(-\ln \lambda_1)]$$
$$= u - (.47588)\beta \quad ,$$

where <u>ln</u> denotes logarithm to the base <u>e</u>. The quantity in brackets may be obtained, for example, by looking up the value $\lambda' = .20$ under the column headed Φ_y in Table 2 of the National Bureau of Standards' forthcoming probability tables for extreme values (reference 4).

Thus $\hat{u}_1'$, which estimates u', not u, requires a knowledge of the other parameter β in order to estimate u without bias. This disadvantage offsets somewhat the greater efficiency of $\hat{u}_1'$ over $\hat{u}_1$. The less efficient estimator $\hat{u}_1$ can be used without knowledge of β. If β is known, then the estimator to use is

(10) $$v_1 = x_{.20n} + .4759\,\beta .$$

This is unbiased and also most efficient among the n order statistics. This result is listed in Table I, columns (1) - (6), line 1.

b. Use of two or three sample values. If instead of one order statistic we are willing to use k suitably chosen ones out of the n, we can estimate the mode by using

(11) $$\hat{u}_k = \frac{1}{k}\sum_{i=1}^{k} x_{n_i} .$$

The generalizations for equations (6), (7), and (9) are, respectively,

(12) $$E(\hat{u}_k) = \frac{1}{k}\sum_{i=1}^{k} u_i \ ,$$

(13) $$\sigma^2(\hat{u}_k) = \frac{1}{n}\Big[\sum_{i=1}^{k} \frac{\lambda_i(1-\lambda_i)}{[f(u_i)]^2} + 2\sum_{i<j} \frac{\lambda_i(1-\lambda_j)}{f(u_i)\,f(u_j)}\Big] ,$$

(14) $$\frac{1}{k}\sum_{i=1}^{k} u_i = u + \Big\{\frac{1}{k}\sum_{i=1}^{k}[-\ln(-\ln \lambda_i)]\beta\Big\} ,$$

where

(15) $$F(u_i) = \lambda_i \ , \qquad i = 1, 2, \ldots, k.$$

For $k = 2$ it is found by trial that

$$\lambda_1 = .08, \qquad \lambda_2 = .40$$

give optimum spacing. After taking account of the bias, given by the term in β in (14), the unbiased estimator to use is

$$v_2 = (1/2)(x_{.08n} + x_{.40n}) + .4074\,\beta \ ,$$

with efficiency close to 82 percent as shown in Table I, columns (1) - (6), line 2.

Likewise for $k = 3$, the optimum estimator is

$$v_3 = (1/3)(x_{.05n} + x_{.20n} + x_{.45n}) + .4494\,\beta \ .$$

For this statistic the efficiency rises to over 88 percent, as shown in Table I, columns (1) - (6), line 3.

Computation rapidly becomes more complicated for higher values of k. If we were to continue till $k = n$ then all n sample values would be drawn into the estimator $\hat{u}_k = \hat{u}_n$. I.e., the estimator in question would then be merely the sample mean corrected for bias,

$$\hat{u}_n = \hat{u} = \bar{x} - C\beta \quad ,$$

with efficiency .6079, as shown in the last line of column (6) of Table I. Hence if we were to compute optimum estimators for increasing values of k, it is noteworthy that efficiency would not always keep increasing. In fact, we must ultimately reach the efficiency .6079 when all n sample values have been taken, which implies that for some value of k, the efficiency would reach its maximum value and from then on, the efficiency would actually

become worse, as more sample values were drawn into the estimator.

Example 1. It is interesting to apply these modal estimators to the data given in Example 1 of reference 1, which uses a different method. The data, consisting of 485 maximum values of gust velocity, one maximum for each of 485 traverses of thunderstorms, are tabulated in Table II taken from that source. For rough comparison purposes, the value $\beta = 1/\alpha = 4.8263$, found in the example of reference 1, is assumed here. The results of estimation by means of the optimum spacing procedures of the present report are:

$$\underline{n = 485}$$

$k = 1$: $v_1 = x_{97} + (.4759)(4.8263) = 10.2258 + 2.2968 = 12.5226$ (assuming that linear interpolation to find the 97th value is valid)

$k = 2$: $v_2 = \frac{1}{2}(x_{38.8} + x_{194}) + (.4074)(4.8263) = \frac{1}{2}(7.7630 + 13.4483) + 1.9662 = 12.5719$

$k = 3$: $v_3 = \frac{1}{3}(x_{24.25} + x_{97} + x_{218.25}) + (.4494)\cdot(4.8263) = \frac{1}{3}(6.4405 + 10.2258 + 14.3000) + 2.1689 = 12.4910$.

These values are listed in Table III and compare very well, considering the saving in computing labor, with

the value u = 12.8370, found in reference 1 by means of the formula $u = \bar{x} - \frac{C}{\alpha}$ (derived by E. J. Gumbel in Lecture 3, page 18 of reference 2), which necessitates calculating the mean of all 485 (grouped) values. The "efficiency" (as here defined) of this latter estimator of u (assuming $1/\alpha = \beta$ is known) given by the Gumbel type of approach can be shown to be 60.8 percent — less even than the 64.8 percent (Table I, column (6)) that can be obtained by using only <u>one</u> ("optimum") sample value.

Table III shows that the "optimum" estimates obtained from the estimators v_1, v_2, v_3, (column (2)), although reasonably close to that given by Gumbel's estimator $\hat{u}$ (column (5)), are all less than $\hat{u}$. Ideally, the various estimates should scatter among each other in random fashion. The fact that they do not may be attributed to grouping (see remark below) or to the fact that the data depart from an underlying extreme-value distribution. It is even possible, though very unlikely, that the sample size n = 485 is not sufficiently large for the normality assumptions made above to be sufficiently valid.

<u>Remark</u>. The closeness of one estimate to another computed from the same data by a different estimator should, of course, be regarded merely as suggestive, rather than as a definite indication of a statistical

property of the estimators. The characteristics of estimators can be studied dependably only by theoretical methods or by extensive experimental sampling. For large sample sizes such as in the above example some weight can be given to the numerical relations obtained, but it should always be kept in mind that these might be upset through sampling variation if another sample of observations were taken.

It should also be noted that since the data are only given in grouped form, interpolation is necessary in order to obtain the desired order statistics. It is likely that more satisfactory estimates would be obtained from data given in full detail. Data should always be available in their original observed form, before grouping or other processing is performed.

The above remark is relevant to all of the examples given below.

2. Equidistant spacing.

Beside optimum spacing, there are two other methods of selecting k statistics in a sample which have been suggested. One of these, the simplest possible, is to divide the ranks 1, 2, ..., n into equal groups, i.e. by taking

$$\lambda_i = \frac{i}{k+1} .$$

Thus, for k = 9, we select the 9 values

$$x_{.1n},\ x_{.2n},\ \ldots,\ x_{.9n},$$

the values of which divide the n ranks into 10 equal groups.

The equidistant method has the advantages that no extensive trial calculations are involved and that it is easier to remember which values to pick out of a sample. Moreover, the work of computing the estimates is as simple as finding quartiles or percentiles.

A disadvantage of the method is that it does not depend upon the form of the population under observation. Exactly the same numerical estimates will be obtained for the normal, Type III, or practically any continuous population. Thus, conceivably we may lose a warning signal which would ordinarily be provided by an estimator sensitive to the form of population, since in that case the estimates would be out of line with experience more frequently if the observed data did not actually come from an extreme-value population than if they did come from it.

Another reason for studying the method of equidistant spacing is that it is intimately related to certain plotting procedures advocated by E. J. Gumbel (reference 1, Lecture 2, pp. 20-21), as will be explained below.

The bias and efficiency of the equidistant method can be obtained by the same procedures as used in optimum spacing, embodied in equations (11) to (15), with $\lambda_i = i/(k+1)$. The results for the first 5 values of k are shown in Table I, columns (7) and (8). The method overestimates the parameter u by substantial amounts, depending upon the parameter β

(assumed known), and a negative correction (column (7)) is necessary in each case, the smallest being $(-.3665\beta)$ for $k = 1$. It can be shown that for increasing k the correction tends toward $(-C\beta)$, where $C = 0.577215\ldots$ is Euler's number.

Efficiency (column (8)) of the equidistant method is relatively low, starting from just under one-half for one order statistic ($k = 1$) and rising to about 63 percent for estimation by five order statistics. This is not even as much as that obtainable from one order statistic under optimum spacing, which is 64.8 percent (column (6)).

The relationship of the equidistant method to Gumbel's plotting method is as follows. Gumbel uses a special probability graph paper (Figure 2) and plots horizontally the values in the sample of extreme observations, $x_1, x_2, \ldots, x_N$, arranged in increasing size. For each x_i he plots the corresponding ratio $i/(N + 1) = \lambda_i$ along a vertical non-uniform probability scale. Parallel to this probability scale is a uniform scale of the function

$$y_i = -\ln(-\ln \lambda_i) \quad . \tag{16}$$

The reasoning then proceeds as follows. If the data come from a true extreme-value distribution, then the points (x_i, y_i) should be situated closely about the straight line

$$x = u + \frac{y}{\alpha} = u + \beta y \quad , \tag{17}$$

where u and β are the parameters of the underlying extreme-value distribution.

This follows from the definition of the extreme-value distribution. The quantity λ, the cumulative probability associated with the observed variable x, is by equation (1) related to x by

$$\lambda = e^{-e^{-\alpha(x-u)}} \quad ;$$

i.e.,

$$\alpha(x-u) = -\ln(-\ln\lambda) \quad .$$

Hence the corresponding y defined by (16) satisfies

$$y = -\ln(-\ln\lambda) = \alpha(x-u) \quad .$$

i.e.,

$$x = u + \frac{y}{\alpha} = u + \beta y \quad ,$$

if x is taken as the variable whose values we wish to predict.

The method given for fitting the straight line (17) is either by eye or by a modification of least squares which allows the line to pass through the mean of the x's, $\bar{x}_N$, and the mean of the corresponding y's, $\bar{y}_N$, computed for the observed sample values. This condition is given by the relation

$$(18) \qquad u = \bar{x}_N - \beta\bar{y}_N$$

connecting the (estimates of the) parameters u, β.

Reference to equation (14) shows that this (sample) mean of the y's, namely,

$$(19) \qquad \bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{N} [-\ln(-\ln\lambda_i)] = b_N \, ,$$

$$\lambda_i = \frac{i}{N+1} \quad ,$$

is precisely the quantity in curly brackets in (14), if N is put for k. The values of this quantity have already been computed in connection with the equidistant method and are given as the coefficients of $(-\beta)$ in Table I, column (7), and designated by b_N. The quantity $b_N = \bar{\bar{y}}_N$ approaches Euler's number, C, as N becomes infinite. The equidistant method is thus essentially equivalent to that of Gumbel's plotting positions, with the difference that sample size N for Gumbel corresponds, not to sample size n for the equidistant method, but to the number k of order statistics selected out of the n.

These considerations make it possible to find the bias in Gumbel's method considered as providing an estimator of the mode u given by equation (18), namely,

$$\hat{u} = \bar{\bar{x}}_N - \beta\bar{\bar{y}}_N = \bar{\bar{x}}_N - \beta b_N \quad .$$

The bias is

$$(20)\quad B_N = E(\hat{u}) - u = E(\bar{\bar{x}}_N) - \beta b_N - u = (u + C\beta) - b_N\beta - u = (C - b_N)\beta,$$

since the expected value of the sample mean is the same as the population mean, namely,

$$(21)\qquad E(x) = u + C\beta \quad .$$

Example 2. Using the same data as in example 1, we obtain the results derived in Table IV. These values are compared with other estimates in Table III. It appears that the estimates given by equal spacing differ from those given by the sample mean by roughly the same order of magnitude (about .2 to .6) as do the estimates given by optimum spacing in Example 1. This time, however, the estimates are all greater than that

given by the sample mean, whereas in Example 1 they were all less. (See also discussion and Remark under Example 1.)

3. Method of expected values.

The last method of spacing which will be presented acts as a compromise, combining some of the features of the other two. This method, based on expected values, avoids the extensive computation of optimum spacing by making use of a special table and, unlike the method of equal spacing, is sensitive to the form of population from which the data are assumed taken.

The method of expected values is as follows. The variate values u_i in (15) are chosen equal to the expected values of the order statistics of a sample of size k from the two-parameter population of largest values. The relation (15) then determines the spacing λ_i.

If $x_{i,k}$ is the i^{th} order statistic in a sample of k from the extreme-value distribution with parameters u, β, then its expected value is

$$E(x_{i,k}) = u + E(y_{i,k})\beta \quad ,$$

where $E(y_{i,k})$ is the expected value for the distribution whose parameters are u = 0, β = 1.[3] The values of $E(y_{i,k})$ have been computed by the National Bureau of Standards in a table (reference 6) prepared at the suggestion of B. F. Kimball, for

[3] This follows from the fact that, by equation (1), if x is the variate in the population $F(x) = \exp[-e^{-\alpha(x-u)}]$, and y the variate in $F_1(y) = \exp(-e^{-y})$, which is the population with parameters u = 0, β = 1, then x and y are related by $x = u + \frac{y}{\alpha} = u + y\beta$.

$i = 1(1)n$ or 25, whichever is smaller, and $n = 1(1)10(5)60$ $(10)100$. From the definition of the expected value method we have

$$u_i = E(x_{i,k}) = u + E(y_{i,k})\beta .$$

For use presently we notice that from this relation and equation (12) we have

$$E(\hat{u}_k) = \sum_{i=1}^{k} \frac{1}{k} u_i = E\left(\sum_{i=1}^{k} \frac{1}{k} x_{i,k}\right) = E(\bar{x}_k),$$

where $\bar{x}_k$ is the ordinary mean of a sample of size k. Since the expected value of the sample mean is always the population mean, we have by equation (21)

$$(22) \qquad E(\hat{u}_k) = E(x) = u + C\beta .$$

Returning to determination of λ_i, we have from equations (1), (15), and the relation $\alpha\beta = 1$,

$$\lambda_i = F(u_i) = \exp[-e^{-\alpha(u_i-u)}] = \exp(-e^{-E(y_{i,k})}).$$

Since $E(y_{i,k})$ are known, the values λ_i can be looked up in a table of the extreme-value cumulative distribution $\exp(-e^{-y})$ such as Table 1 of reference 4.

Determination of the λ_i for the expected value method is shown in Table V, and the biases and efficiencies of the method, in Table I, columns (9) and (10). If β is known, then the method, for practical purposes, is unbiased, since the bias, $C\beta$, and the correction, $-C\beta = -0.5772\beta$, are always the same known values for any k. This follows easily from equation (22), since the bias of $\hat{u}_k$ in estimating u is, by definition,

$$E(\hat{u}_k) - u = C\beta$$

The efficiencies of the expected values method are even lower than for equidistant spacing, starting from about 42 percent for estimation by a single order statistic and being only 57 percent for estimation by five statistics.

The method of expected values bears the same relation to a graphical method advocated by Kimball as does the equidistant method to the plotting procedure of Gumbel. Kimball's method consists in modifying Gumbel's graphical procedure by simply using for λ_i the values given by the method of expected values instead of the simple ratios $i/(k + 1)$, and leaving everything else unchanged. The result is to replace equation (19) by

$$b_N' = \frac{1}{N} \sum_{i=1}^{N} y_i = \frac{1}{N} \sum_{i=1}^{N} E(y_{i,N}) \text{ ,}$$

which has been shown equal to C. Hence the bias in Kimball's method is

$$B_N' = (C - C)\beta = 0 \text{ ;}$$

i.e., his method has the theoretical merit of giving an unbiased position for the centroid of the fitted line.

Example 3. Again using the data of Example 1, we perform the calculations for the expected values method as in Table IV, the results of which are listed in Table III for comparison. It appears that this method gives estimates about 10 percent closer to those given by the sample mean than the equidistant method gives. Again, the values are all higher than for the

sample mean. (Compare Example 1.)

If one of these two systematic methods of spacing order statistics is used, it appears that at least 3 should be taken, since the last two examples indicate that k = 1 and 2 give relatively widely discrepant values as compared with larger values of k. This precaution, however, does not seem to be needed with optimum spacing.

VI. ESTIMATION OF DISPERSION PARAMETER β

1. <u>Use of two sample values.</u>

Mosteller (reference 3) has discussed the use of the quasi-range

$$\hat{\beta}_2 = (x_{n_2} - x_{n_1})/c$$

to estimate the standard deviation of a normal population from a sample of n. This estimator may also be used in the extreme-value case for estimating β, provided <u>c</u> is chosen as

$$c = u_2 - u_1 \quad ,$$

where u_1, u_2 are defined by

$$F(u_i) = \lambda_i = n_i/n \ , \quad i = 1, 2,$$

as in equation (15).

Again using trial methods, we find the optimum spacing to be given by

$$\lambda_1 = .03, \quad \lambda_2 = .85 \quad ,$$

with c = 3.07159. Hence for two sample values, the optimum estimator for β is

$$\hat{\beta}_2 = 0.3256(x_{.85n} - x_{.03n}) \quad , \tag{23}$$

and its efficiency [4] (as defined earlier) is 59.5 percent. Two points are worth noting about $\hat{\beta}_2$. First, it is unbiased, bias being taken care of by the numerical factor. Secondly, unlike the estimator for u, the estimator for β does not depend on a knowledge of the other parameter u.

2. Use of four sample values.

The above result for two sample values can be improved by taking two additional values. While exact determination has not been carried out, a considerable amount of trial indicates that probably the best efficiency for four values is reached by using the estimator

$$\hat{\beta}_4 = 0.2026\,(x_{.85n} + x_{.70n} - x_{.10n} - x_{.03n}) .$$

Its efficiency is 68.9 percent. As for the case of two values, the numerical factor eliminates the bias, and use of the estimator does not require a knowledge of the other parameter u.

Example 4. If we use the same data as in Example 1, the above two estimators of β give the following results:

2 values: $\hat{\beta}_2 = 0.3256(x_{412.25} - x_{14.55})$

$= 0.3256\,(21.4583 - 3.9182) = 5.7111$

4 values: $\hat{\beta}_4 = 0.2026(x_{412.25} + x_{339.5} - x_{48.5} - x_{14.55})$

$= 0.2026(21.4583 + 18.4754 - 8.2708 - 3.9182)$

$= 5.6211$

[4] In the case where both parameters are unknown, the concept of efficiency defined earlier for the one-parameter case is not strictly applicable. See section VII below for further comments.

These values do not seem to be in as good agreement with the value $\beta = 1/\alpha = 4.8263$, found in reference 1 (by means of the formula $1/\alpha = \frac{\sqrt{6}}{\pi} s$ given by Gumbel) as is the agreement between the order-statistics estimates of u and the value of u found in reference 1 by use of the sample mean. This is not too surprising in view of the much greater reduction of computing work through avoiding calculation of the standard deviation as compared with the reduction by avoiding the mean, since ordinarily it is to be expected that more information is lost when the reduction in labor is greater, other things being equal.

For comparison, the efficiency of the Gumbel estimator

$$\hat{\beta} = \frac{\sqrt{6}}{\pi} s$$

which is (asymptotically) unbiased, is approximately 39 percent. This value is obtained on the assumption that the sample size n = 485 is large enough so that the first one or two terms in a Taylor series expansion of s furnish a good approximation. This assumption has been found to give usable results in similar situations. Determination of its validity depends upon the exact evaluation of the variance of s, which requires a prohibitive amount of numerical integration by ordinary methods. [5]

---

[5] Special sampling procedures in process of development give promise of shortening the labor to within feasible limits.

VII. ESTIMATION OF MODE WHEN BOTH PARAMETERS ARE UNKNOWN

When both parameters are unknown, one might seek to try a combination of the above methods. One would first estimate β by the quasi-range with two or four order statistics and then use this value in conjunction with one of the estimators for u discussed previously. A complete discussion of this method would, however, involve the question of joint estimation of two unknown parameters and is outside the scope of this report. However, one simple case will be touched on briefly.

We can try a combination of the most efficient single sample value with a multiple of the difference of two of the sample values. The estimator is, from (10) and (23),

$$\hat{u}' = x_{.20n} + .47588\hat{\beta}_2$$

$$= x_{.20n} + .15493(x_{.85n} - x_{.03n}) \tag{24}$$

This estimator is unbiased since $\hat{\beta}_2$ is unbiased. Its variance is, by a modification of equation (13),

$$V(\hat{u}') = \Big\{ \frac{\lambda_2(1-\lambda_2)}{f_2^2} + (.15493)^2\Big[\frac{\lambda_3(1-\lambda_3)}{f_3^2} + \frac{\lambda_1(1-\lambda_1)}{f_1^2}\Big]$$

$$+ 2(.15493)\Big[\frac{\lambda_2(1-\lambda_3)}{f_2 f_3} - \frac{\lambda_1(1-\lambda_2)}{f_1 f_2}\Big]$$

$$- 2(.15493)^2 \frac{\lambda_1(1-\lambda_3)}{f_1 f_3} \Big\} \frac{\beta^2}{n} ,$$

where $\lambda_1 = .03$, $\lambda_2 = .20$, $\lambda_3 = .85$, and the other quantities are as defined in (13) and (15) above. When evaluated, this gives

$$V(\hat{u}) = 1.7423\beta^2/n .$$

The corresponding Gumbel estimator of u, when β is unknown, is

$$\hat{u}_G = \bar{x} - \frac{\sqrt{6}}{\pi} Cs ,$$

where s is the sample standard deviation. As in the case of $\hat{\beta}$, the Gumbel estimator of β, the exact calculation of the variance $\hat{u}_G$ may involve prohibitive labor. The approximation method used in connection with $\hat{\beta}$ gives

$$V(\hat{u}_G) = 1.4042\beta^2/n .$$

The concept of efficiency previously used, where only one parameter was unknown, is not directly applicable to the case of more than one unknown parameter, but involves the concept of joint estimation mentioned above. Instead, we may proceed as follows.

In the one-parameter case, two different unbiased estimators $\hat{u}_1$, $\hat{u}_2$, of one of the parameters, say u, were compared by dividing the variance of each into a common value, namely, the theoretically smallest value (under certain conditions) which the variance of any (unbiased) estimator of u could have. This furnished for each estimator an index, called efficiency, of the form

$$\text{eff}(\hat{u}_i) = \frac{\sigma^2(\hat{u})_{min}}{\sigma^2(\hat{u}_i)} , \quad i = 1, 2.$$

Instead of evaluating each fraction separately, we could compare the efficiencies of $\hat{u}_1$ and $\hat{u}_2$ by simply taking their ratio,

$$\frac{\text{eff }(\hat{u}_1)}{\text{eff }(\hat{u}_2)} = \frac{\sigma^2(\hat{u}_2)}{\sigma^2(\hat{u}_1)} = k, \text{ say, } \quad 0 \leq k \leq 1,$$

and say that the estimator $\hat{u}_1$ is k times as efficient as the estimator $\hat{u}_2$.

In the two-parameter case we can also compare estimators, in similar fashion, by taking the ratio of their variances, as k above, and designating this ratio the "relative efficiency" of $\hat{u}_1$ relative to $\hat{u}_2$, with the understanding that the word "efficiency" will not necessarily have the same meaning as in the one-parameter case.

We then have, subject to the approximations described above in connection with the estimator $\hat{\beta}$,

$$k = \frac{V(\hat{u}_G)}{V(\hat{u}')} = 0.8060 \quad .$$

That is, the estimator of u constructed from order statistics is about 81 percent as "efficient" as the estimator given by Gumbel, in the case where both parameters are unknown. Thus the order statistic estimator $\hat{u}'$ can be used for rapid calculation without too great a loss of accuracy over that of the estimator of Gumbel which requires computation of sums of squares.

This discussion is for β unknown. If β is known, then $\hat{u}'$ should be replaced by one of the more efficient estimators listed in Table I.

Example 5. Applying the estimator in (24) to the same data as used before with β now unknown gives, from the results of examples 1 and 4 above,

$$\hat{u}' = 10.2258 + .47588\ (5.7111)$$
$$= 12.9436$$
$$\hat{\beta}_2 = 5.7111\,.$$

The estimate of u is remarkably close to that obtained by Gumbel's formula, 12.8370, but does not require the labor of squaring, adding, and taking a square root. The estimate of β is of course the same as in example 4.

VIII. CONCLUSIONS

Several large-sample unbiased estimators of the parameters of an extreme-value distribution have been developed which, with one exception, appear to have greater efficiency than those derived from methods of E. J. Gumbel and B. F. Kimball and yet require much less effort in computation, involving essentially a mechanical sorting to find pre-designated ranked values in a sample of size n.

The unbiased estimators which can be recommended are the following linear functions of order statistics, where, for example, $x_{.03n}$ means the $(\frac{3n}{100})^{th}$ observation in the sample when all are arranged in ascending order, and "est. u" means "estimator of u," the subscripts being omitted for simplicity. Application to a sample of n = 485 maximum gust-velocity

observations analyzed in NACA TN 1926 (reference 1) has been found to give satisfactory results.

1. <u>For estimating the mode u when $\beta = 1/\alpha$ is known.</u> Use:

(a) $\text{est.}u = x_{.20n} + .4759\,\beta.$ Efficiency = 65%

(b) $\text{est.}u = \frac{1}{2}(x_{.08n} + x_{.40n}) + .4074\beta.$ Efficiency = 82%

(c) $\text{est.}u = \frac{1}{3}(x_{.05n} + x_{.20n} + x_{.45n}) + .4494\beta.$ Efficiency = 88½%

By comparison, estimation by using the mean of all n values, as in the Gumbel method, is about 61 percent efficient.

2. <u>For estimating u when β is unknown.</u> Use:

$$\text{est.}\ u = x_{.20n} + .1549\,(x_{.85n} - x_{.03n}).$$

This estimator is 81 percent as "efficient" as the estimator given by Gumbel which involves the sample mean and standard deviation, but avoids the work of squaring and summing.

3. <u>For estimating β whether or not u is known.</u> Use:

(a) $\text{est.}\beta = .3256\,(x_{.85n} - x_{.03n}).$ Efficiency = 59½%

(b) $\text{est.}\beta = .2026\,(x_{.85n} + x_{.70n} - x_{.10n} - x_{.03n}).$ Efficiency = 69%

These agree less well with the estimator requiring the standard deviation, but the efficiency of the latter is indicated to be about 39 percent.

Other methods of choosing the order statistics which were examined may have some theoretical advantage and perhaps be somewhat simpler, but they have substantially lower efficiency than those recommended.

Finally, it should be urged that when data are presented for analysis, they should be given in their original detailed form, before grouping or other processing. This would obviate the need for interpolation or other devices which may tend to vitiate the accuracy of the analysis.

Statistical Engineering Laboratory
National Bureau of Standards
Washington, D. C.

15 August 1951

REFERENCES

1. Press, Harry: The Application of the Statistical Theory of Extreme Values to Gust-Load Problems. NACA TN 1926, 1949.

2. Gumbel, E. J.: Lectures on the Statistical Theory of Extreme Values and their Practical Applications. Manuscript to be published by the National Bureau of Standards.

3. Mosteller, F.: On Some Useful "Inefficient" Statistics. The Annals of Mathematical Statistics, vol. XVII, no. 4, Dec. 1946, pp. 377-408.

4. National Bureau of Standards: Probability Tables for the Analysis of Extreme-Value Data. (In press.)

5. Kimball, B. F.: Assignment of Frequencies to a Completely Ordered Set of Sample Data. Trans., Am. Geophys. Union, vol. 27, no. 6, 1946, pp. 843-846.

6. Table of the First Moment of Ranked Extremes. (Report to be submitted to NACA by the National Bureau of Standards.)

7. Cramér, Harald: Mathematical Methods of Statistics. Princeton University Press (Princeton), 1946, p. 480, equation (32.3.3a).

TABLE I

Comparison of three types of methods of estimating the mode u of the distribution of largest values: (i) use of k order statistics spaced in three different ways, β $(= \frac{1}{\alpha})$ known; (ii) use of sample mean, β known; (iii) use of "optimum" order statistic and quasi-range, β unknown.

	(i) Three methods of spacing k order statistics (β known)									(ii) Sample mean (β known) $\hat{u} = \bar{x} - C\beta$		(iii) Order statistic and quasi-range (β unknown) $\hat{u}' = x_{.20n} + .15493 \cdot (x_{.85n} - x_{.03n})$	
	Optimum Spacing					Equidistant Spacing $\lambda_i = \frac{i}{k+1}$		Expected Values (λ_i are given in Table V)					
k	λ_1	λ_2	λ_3	Bias correction (add to estimate)	Efficiency = $\frac{\sigma^2(\text{most eff.})}{\sigma^2(\text{estimator})}$	Bias correction to estimate, $-b_k\beta$	Efficiency	Bias correction = $-C\beta$ = $-.5772\beta$	Efficiency	Bias	Efficiency	Bias	Relative Efficiency
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	.20			0.4759β	0.6476	-0.3665β	0.4804	-0.5772β	0.4188	(For all sample sizes n)			
2	.08	.40		.4074β	.8188	- .4043β	.5761	- .5772β	.5108				
3	.05	.20	.45	.4494β	.8857	- .4286β	.6095	- .5772β	.5505				
4						- .4458β	.6241	- .5772β	.5651				
5						- .4588β	.6313	- .5772β	.5746				
⋮													
n(very large)				$C\beta$ = .5772β	$\frac{6}{\pi^2}$ = .6079	- .5772β	.6079	- .5772β	.6079	(none)	0.6079	(none)	80.6 percent as "efficient" as $\hat{u}_G = \bar{x} - \frac{\sqrt{6}}{\pi} Cs$

TABLE II

Distribution of n = 485 maximum gust velocities

Class interval in ft./sec.	Frequency	Cumulative frequency from smallest
2 to 4	4	4
4 to 6	11	15
6 to 8	27	42
8 to 10	48	90
10 to 12	62	152
12 to 14	58	210
14 to 16	55	265
16 to 18	60	325
18 to 20	61	386
20 to 22	36	422
22 to 24	17	439
24 to 26	18	457
26 to 28	8	465
28 to 30	7	472
30 to 32	6	478
32 to 34	3	481
34 to 36	1	482
36 to 38	2	484
38 to 40	1	485
	485	

Source: Table 2 of reference 1
(1946 Thunderstorm Project data)

TABLE III

Estimates of mode u obtained by various methods

from a sample of n = 485 extremes

$\beta = 4.8263$

k	v_k for three methods of spacing k order statistics			Estimation by sample mean
	"Optimum"	Equidistant	Expected values	$\hat{u} = \bar{x} - C\beta$
(1)	(2)	(3)	(4)	(5)
1	12.5226	13.4130	13.6008	
2	12.5719	13.1876	12.9896	
3	12.4910	13.0767	12.9594	
4		13.0114	12.9008	
5		13.0031	12.9794	
n = 485				12.8370

Source of basic data: NACA TN 1926, Table II

TABLE IV

Calculations for estimating mode u from sample of n = 485 extremes

by use of equal-spacing and expected-values methods with 1 to 5 order statistics

$\beta = 4.8263$

k	Equal spacing						Expected values				
	$\lambda_i = \frac{i}{k+1}$	Rank $n_i = \lambda_i n$	Sample values, x_{n_i} (Interpolated)	Average, $\hat{u}_k = \frac{1}{k}\sum_{i=1}^{k} x_{n_i}$	Bias correction to estimate, $-b_k\beta$ (from Table I, column (7)	Estimate of mode	λ_i	Rank $n_i = \lambda_i n$	Sample values, x_{n_i} (Interpolated)	Average, $\hat{u}_k$	Estimate of mode u, $\hat{u}_k' = \hat{u}_k - C\beta = \hat{u}_k - 2.7857$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	1/2	242 1/2	15.1818	15.1818	-1.7688	13.4130	0.5703	276.5955	16.3865	16.3865	13.6008
2	1/3	161 2/3	12.3333				0.3253	157.7705	12.1990		
	2/3	323 1/3	17.9444	15.1389	-1.9513	13.1876	0.7551	366.2235	19.3516	15.7753	12.9896
3	1/4	121 1/4	11.0081				0.2238	108.5430	10.5982		
	1/2	242 1/2	15.1818				0.5316	257.8260	15.7391		
	3/4	363 3/4	19.2459	15.1453	-2.0686	13.0767	0.8292	402.1620	20.8979	15.7451	12.9594
4	1/5	97	10.2258				0.1696	82.2560	9.6773		
	2/5	194	13.4483				0.4068	197.2980	13.5620		
	3/5	291	16.8667				0.6416	311.1760	17.5392		
	4/5	388	20.1111	15.1630	-2.1516	13.0114	0.8689	421.4165	21.9676	15.6865	12.9008
5	1/6	80 5/6	9.6181				0.1389	67.3665	9.0569		
	1/3	161 2/3	12.3333				0.3286	159.3710	12.2542		
	1/2	242 1/2	15.1818				0.5202	252.2970	15.5381		
	2/3	323 1/3	17.9444				0.7097	344.2045	18.6297		
	5/6	404 1/6	21.0093	15.2174	-2.2143	13.0031	0.8937	433.4445	23.3464	15.7651	12.9794

NOTE: For comparison, the estimate of u given by the sample mean is $\hat{u} = \bar{x} - C\beta = 12.8370$

TABLE V

Calculation of spacing constants λ_i
for expected values method

k	$E(y_{i,k})$				
	i = 1	i = 2	i = 3	i = 4	i = 5
1	0.5772				
2	-0.1159	1.2704			
3	-0.4036	0.4594	1.6758		
4	-0.5735	0.1061	0.8128	1.9635	
5	-0.6902	-0.1069	0.4256	1.0709	2.1867

k	$\lambda_i = \exp[-e^{-E(y_{i,k})}]$				
	i = 1	i = 2	i = 3	i = 4	i = 5
1	0.5703				
2	0.3253	0.7551			
3	0.2238	0.5316	0.8292		
4	0.1696	0.4068	0.6416	0.8689	
5	0.1389	0.3286	0.5202	0.7097	0.8937

Source: See text, Section V, 3

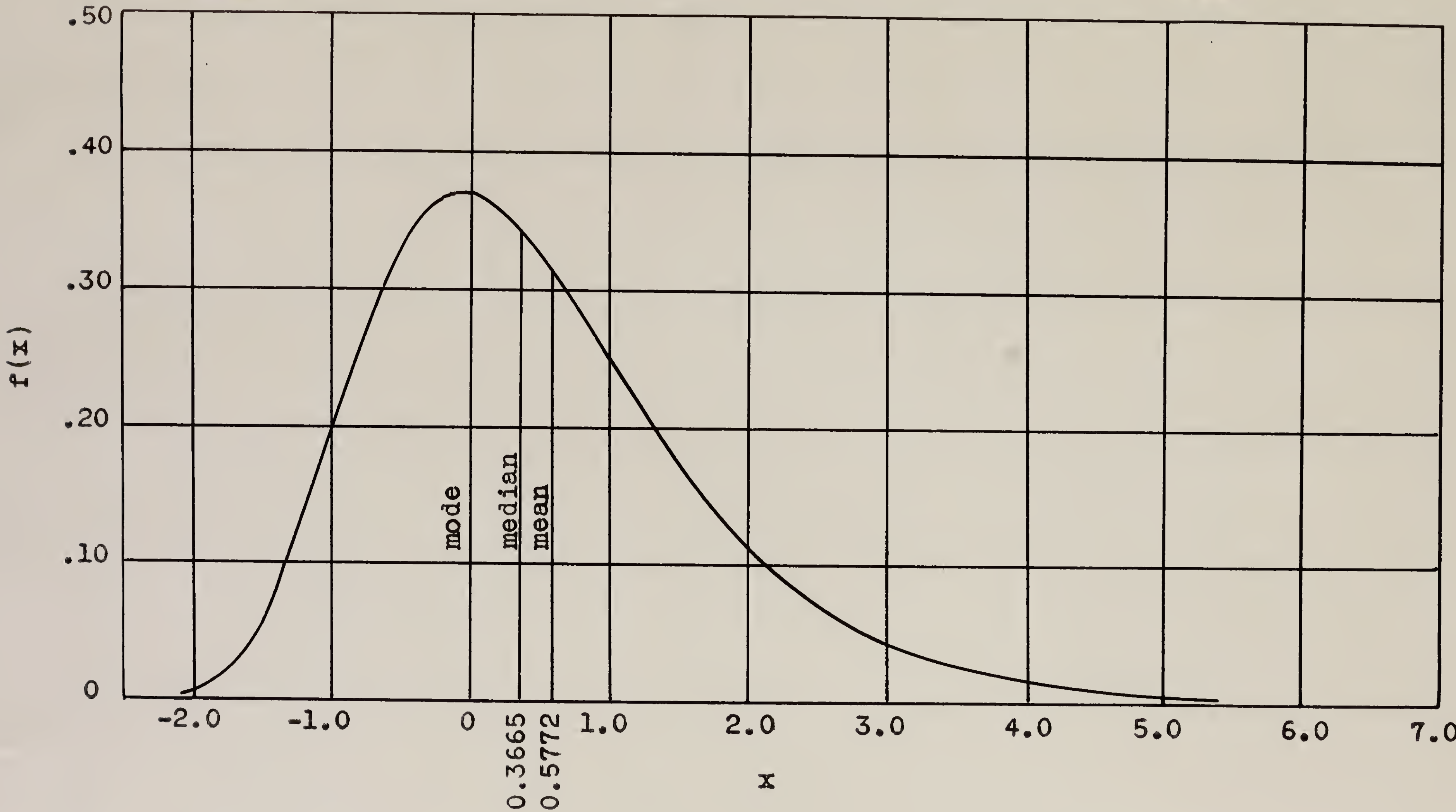

Figure 1 - Density function for extreme-value distribution

with parameters $\alpha = 1$, $u = 0$: $f(x) = e^{-x-e^{-x}}$.

(Adapted from NACA TN 1926, Figure 2)

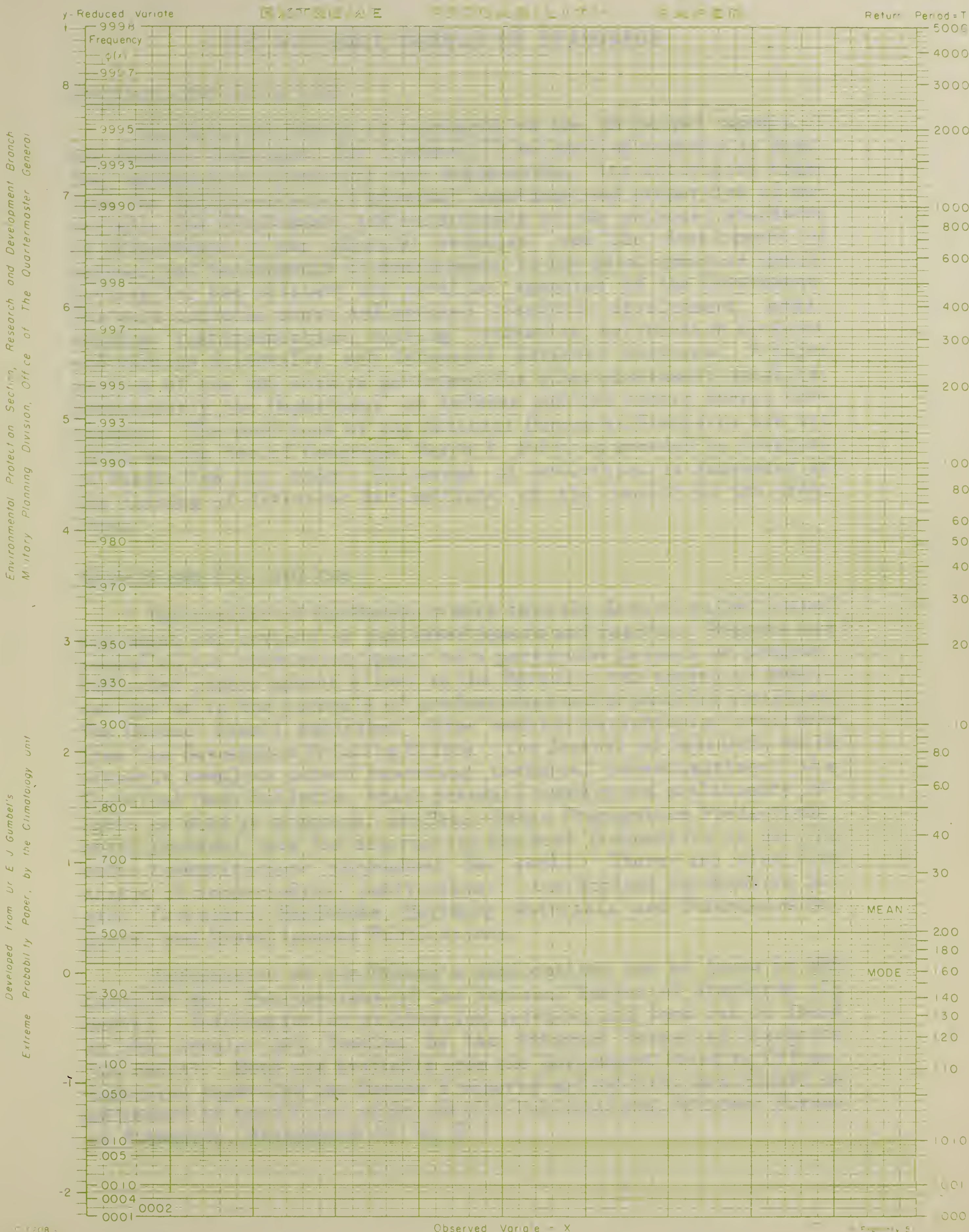

Figure 2. Specimen of extreme-value probability graph paper.

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The National Bureau of Standards is the principal agency of the Federal Government for fundamental and applied research in physics, mathematics, chemistry, and engineering. Its activities range from the determination of physical constants and properties of materials, the development and maintenance of the national standards of measurement in the physical sciences, and the development of methods and instruments of measurement, to the development of special devices for the military and civilian agencies of the Government. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various scientific and technical advisory services. A major portion of the NBS work is performed for other government agencies, particularly the Department of Defense and the Atomic Energy Commission. The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. The scope of activities is suggested in the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: the Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: the Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards (75 cents). Information on calibration services and fees can be found in NBS Circular 483, Testing by the National Bureau of Standards (25 cents). Both are available from the Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.

NBS